

*Roles and Responsibilities  
for Evaluation in Foreign  
Language Programs*



John M. Norris & Yukiko Watanabe

University of Hawai'i at Mānoa  
Department of Second Language Studies  
*Foreign Language Program Evaluation Project*

# Roles and Responsibilities for Evaluation in Foreign Language Programs

## I. Introduction

Increasingly, language teachers are faced with the task of making evaluation happen in their language classrooms and programs, or they are confronted with the realities of evaluations being done from the outside. Given the variety of roles—both positive and negative—that evaluation may play, it is critical that teachers, as well as other participants in language programs, understand their responsibilities in the evaluation process. This booklet offers a procedural guide for helping language educators understand, implement, and use evaluation; it also highlights evaluators' roles, responsibilities, and ethics throughout the evaluation process.

Throughout this booklet, the following working definition of evaluation is used.

**Evaluation:** “Evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming.” (Patton, 1997, p. 23)

## II. The changing landscape of language program evaluation

### PAST

- **JJJOE model:** Short-term, external, cost-effectiveness judgment to terminate or continue a program.
- **Accountability model** (e.g., NCLB): Policy-makers and testers' definition of educational effectiveness is imposed; measurable outcomes.
- **Administrative-managerial model:** To keep everyone in line and satisfy the public.

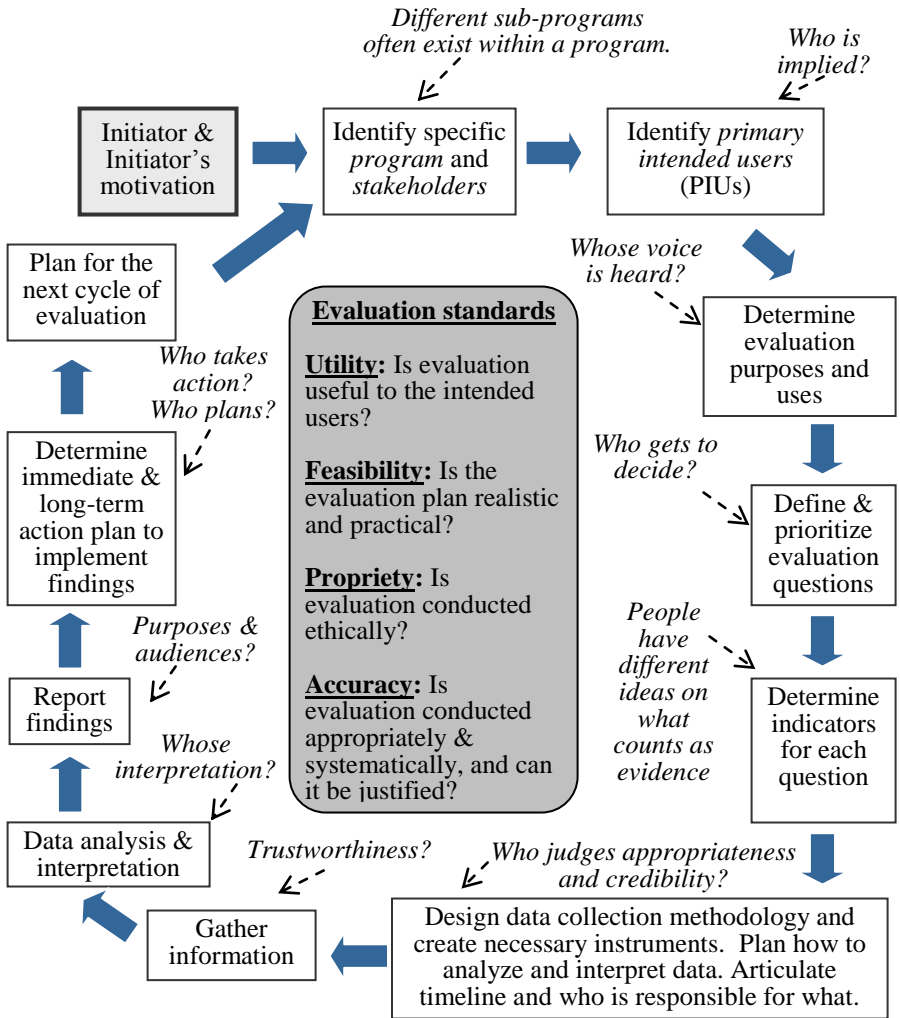
### MORE RECENT

- **Participatory models** (e.g., utilization-focused evaluation): Use evaluation findings for variety of intentional, contextualized ends.
- **Professional responsibility** (e.g., NCATE, TESOL, ACTFL standards): To improve teaching practice for creating effective learning environment.
- **Empowerment evaluation:** Build internal capacity to sustain ongoing, bottom-up evaluation.

- From externally mandated evaluation to internally-motivated evaluation.
- From judgmental uses & testing to multiple uses & methods.
- Participatory process, ownership, and usefulness are promoted.



### III. Overview of use-driven & participatory evaluation process



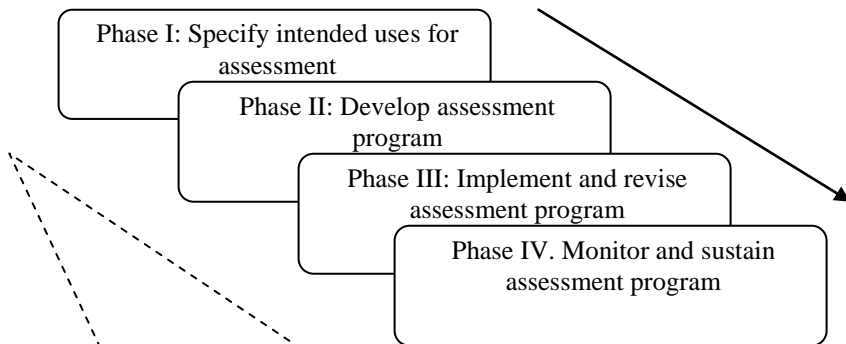
**Stakeholders are:** \_\_\_\_\_

**Primary intended users are:** \_\_\_\_\_

There are various steps involved in planning, implementing, analyzing, interpreting, reporting, and using evaluation. In order to assure usefulness of evaluation, the starting point for evaluation planning should be “for whom” and “why,” rather than “what” and “how” questions.



# Evaluation of a college German FL assessment program



## Who took what responsibility during evaluation?

### *Step 0: Initiator*

→ German department faculty invite evaluation consultant to advise/participate in curriculum, instruction, assessment innovation project.

### *Step 1: Specify program and stakeholders*

→ Consultant conducts context analysis, surveys stakeholders (students, faculty), reviews existing assessment practices, identifies areas of need.

### *Step 2: Identify primary intended users (PIUs)*

→ Small group of decision makers identified (chair, curriculum developer, curriculum coordinator); full faculty endorse user group.

### *Step 3: Determine evaluation purposes and uses*

→ PIU group prioritize uses for evaluation (developing and improving assessment program); faculty and instructors review, revise, authorize.

### *Step 4: Define & prioritize evaluation questions*

→ Faculty and instructors pose general questions about assessment, PIUs operationalize as specific questions for investigation.

### *Step 5: Determine indicators for each question*

→ Consultant elicits types of evidence considered relevant for answering questions from PIUs

### *Step 6: Design data collection methodology*

→ Consultant and PIUs develop methods for gathering data in accord with indicators; faculty and instructors review and advise on feasibility.

### *Step 7: Gather information*

→ Faculty and instructors gather information from students (e.g., test performances, questionnaires); consultant collects additional data

### *Step 8: Data analysis & interpretation*

→ Consultant makes first pass at summarizing and displaying findings; faculty and instructors review findings in meetings, decide on interpretations.

### *Step 9: Report findings*

→ Consultant reports findings in iterations, primarily in full-department meetings, also in short formal reports, letters to stakeholders; PIUs endorse.

### *Step 10: Determine action plan*

→ Full-department review leads to concrete implications and recommendations for improving assessments; PIUs develop new policies and practices

### *Step 11: Plan for the next cycle of evaluation*

→ From one stage to the next, PIUs consider new priorities and propose to the faculty; overall, full-department consideration of findings leads to planning.



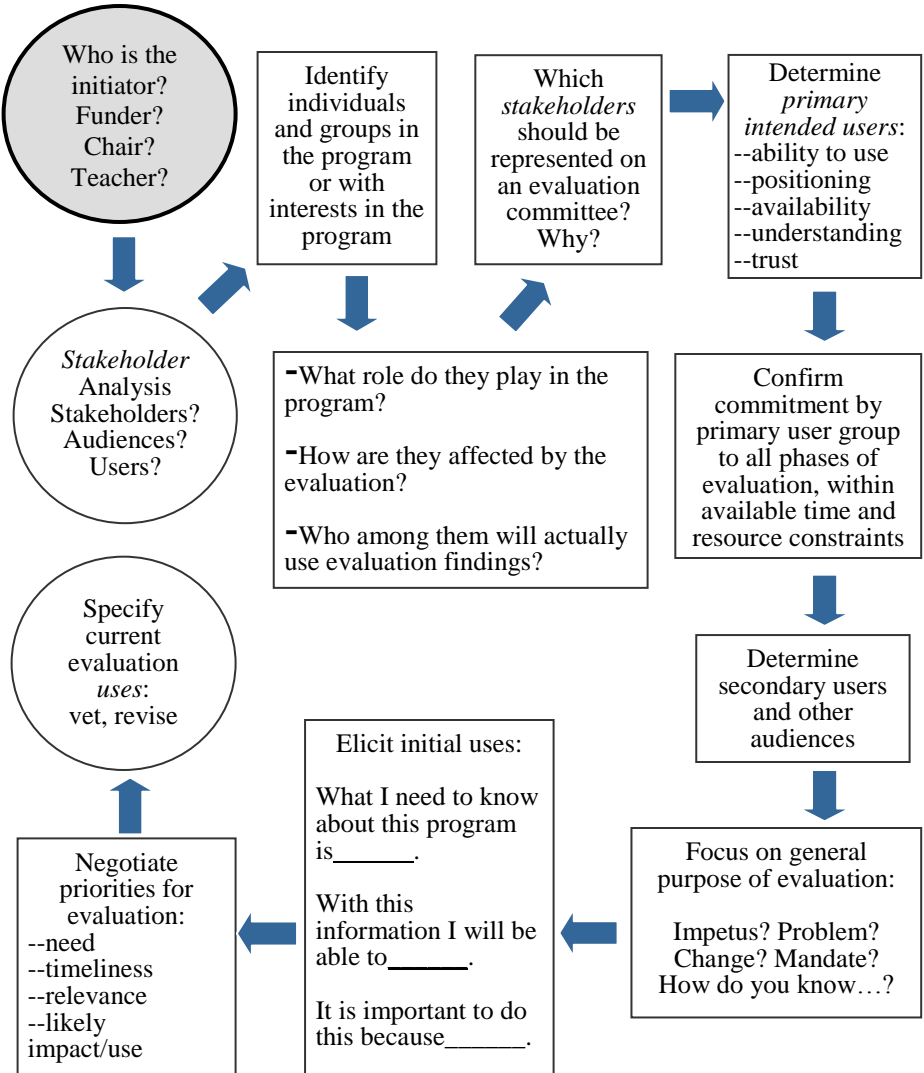
## IV. Language program evaluation examples

Author	Program context	Primary Intended Users	Evaluation focus	Actual evaluation use
Harris (forth-coming)	Nation-wide Irish language education in primary schools in Ireland; three decades of program evaluation	<ul style="list-style-type: none"> <li>- Policy makers (Irish Department of Education and Science, National Council for Curriculum and Assessment)</li> <li>- Primary school teachers</li> <li>- Public (parents, students, media)</li> </ul>	<ul style="list-style-type: none"> <li>- Language achievement outcomes for distinct types and regions</li> <li>- Causal factors underlying achievement differences</li> <li>- Impact of evaluation findings on stakeholders</li> </ul>	<ul style="list-style-type: none"> <li>- Demonstrate some achievement</li> <li>- Explain apparent low achievement based on demographic and implementation factors</li> <li>- Mediate reactions to Irish language education through careful reporting of full evidence</li> </ul>
Llosa & Slayton (forth-coming)	School-district wide kindergarten and first-grade computer-based early reading program	<ul style="list-style-type: none"> <li>- Los Angeles Unified School District</li> <li>- Teachers</li> <li>- Program publisher</li> </ul>	<ul style="list-style-type: none"> <li>- To what extent is the program being implemented?</li> <li>- To what extent does it help English-language learners?</li> <li>- Should it be retained?</li> </ul>	<ul style="list-style-type: none"> <li>- Improve implementation</li> <li>- Adapt instructional schedules</li> <li>- Provide teachers with training, guidelines</li> <li>- Target learners in need of assistance</li> </ul>
Norris (in press)	Undergraduate German program at Georgetown University, Assessment program	<ul style="list-style-type: none"> <li>- Program chair</li> <li>- Curriculum developer</li> <li>- Curriculum coordinator</li> <li>- Faculty</li> <li>- Instructors (graduate TAs)</li> <li>- FL programs across the U.S.</li> </ul>	<ul style="list-style-type: none"> <li>- Understand intended uses of assessment at all levels, from individuals to classes to program</li> <li>- Develop assessment policies, instruments, and practices to meet uses</li> <li>- Investigate effectiveness of assessments to revise and improve</li> </ul>	<ul style="list-style-type: none"> <li>- Develop assessment policies to align assessment and curriculum.</li> <li>- Create curriculum-based assessment methods (placement in particular).</li> <li>- Monitor and revise assessment practices</li> <li>- Generate awareness and sustained attention for assessment program</li> </ul>
Fudano (2005)	A ten-week summer intensive Japanese program at Kanazawa Institute of Technology	<ul style="list-style-type: none"> <li>- Program manager (also the evaluator)</li> <li>- Chancellor and the board of administration of the university</li> </ul>	<ul style="list-style-type: none"> <li>- To what extent did the KIT-IJST program achieve its intended goals?</li> <li>- Satisfaction of program participants and cooperating personnel with the program outcomes</li> <li>- Appropriateness of program content and management</li> </ul>	<ul style="list-style-type: none"> <li>- Change in pre-departure testing</li> <li>- Change in curricular content</li> <li>- Illuminated program mission and goals perceived by univ upper admin</li> <li>- Demonstrated program value to univ upper admin.</li> <li>- No funding cuts</li> </ul>



## A. Intended use and user identification: *Who and why?*

Various purposes can guide program evaluation, such as to develop and improve a program, generate knowledge and understanding about a program, determine program effectiveness, and empower program stakeholders. In order to achieve meaningful evaluations that actually get used, evaluation purpose and use should be generated by the intended users of evaluation. This diagram depicts a participatory process for identifying intended evaluation uses by intended users.



### *A reflection question for the evaluator:*

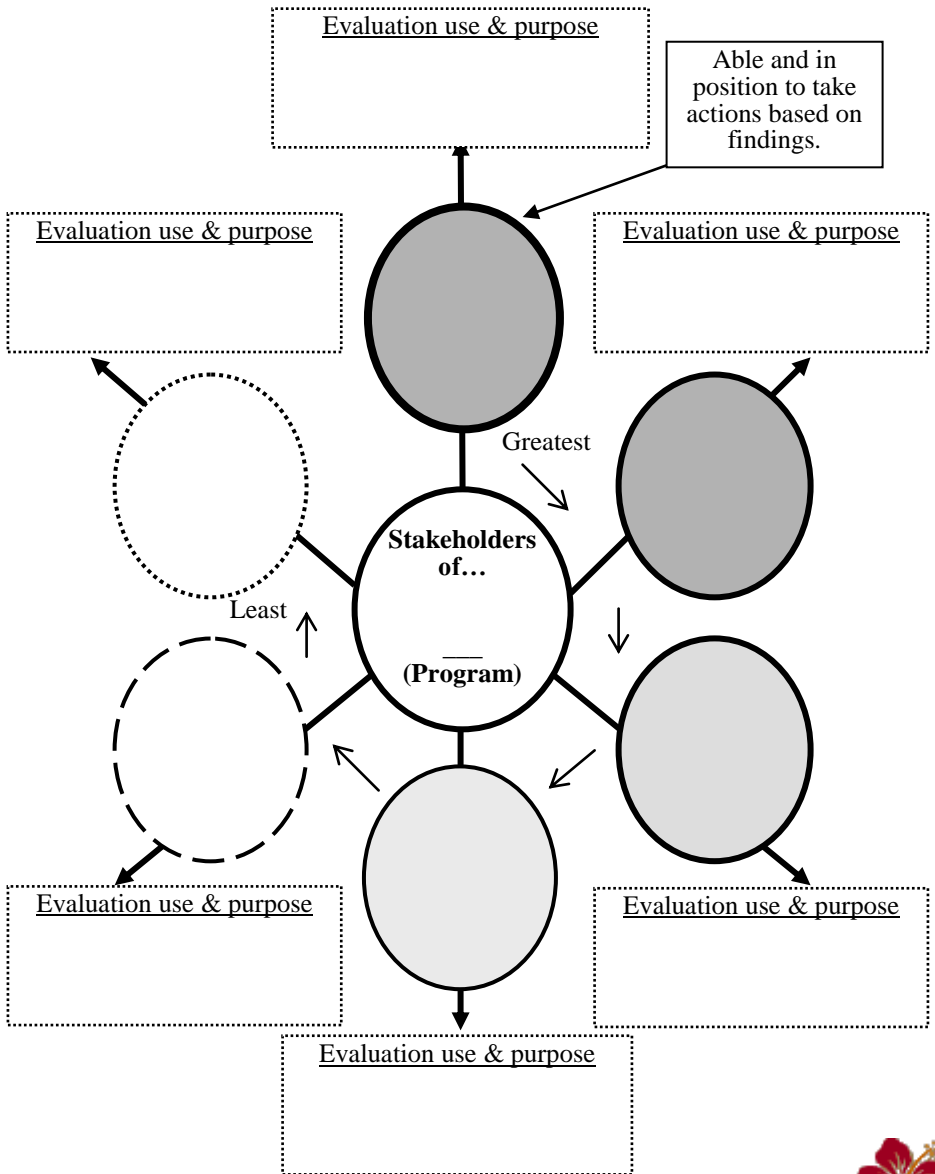
As an initiator of evaluation, how might you seek participation from the stakeholders who make decisions and are invested/interested in the program?



**Task 1:**

- ❖ Identify the internal and external stakeholders of your program and list them clockwise according to the degree (greater to lesser) of interest in evaluation and the impact they can make on the program based on the evaluation findings.
- ❖ Within the stakeholders, who is the initiator? What is the reason (impetus) for the initiator to conduct evaluation?
- ❖ Can you specify an evaluation use and purpose for each stakeholder group?

**Initiator's impetus:** \_\_\_\_\_



## B. Evaluation questions and methods: *What* and *how*?

There are advantages and disadvantages to any data collection methodology. Primary intended users need to make an informed decision—based on the intended uses of an evaluation—about which methodology is most appropriate, feasible, cost-effective, credible, and realistic for getting the information they need. Before making methods decisions, it is important to have a good idea of the questions being asked and the types of indicators (phenomena that will be useful in answering the questions) that will be observed.

**Task 2:** For each evaluation question *intended users* may have, list the kind of indicators that need to be observed, the *key informants*, and the best timing to collect data. See the first example.

### ***Evaluation question 1:***

What aspects of language teaching do the new teachers most need to improve in?

→ **Context:** The first-year teacher development program for Japanese junior high school teachers of English.

→ **Use:** Develop teacher induction program content that will best meet teachers' needs.

### **Indicators:**

(a) First year teachers' immediate concerns they are facing in daily teaching; (b) the difficulties 2nd year teachers have faced in their first year; (c) immediate feedback for improvement from the first year teachers' in-service training mentors.

**Key informants:** First and second year teachers, in-service mentors

**Timing:** At the end of the first year, survey or interview the cohort on their first-year experience. Monitor 1<sup>st</sup> yr teachers and their mentors at the beginning and throughout the year for their immediate needs.

### ***Evaluation question:***

### **Indicators:**

**Key informants:**

**Timing:**

### ***Evaluation question:***

### **Indicators:**

**Key informants:**

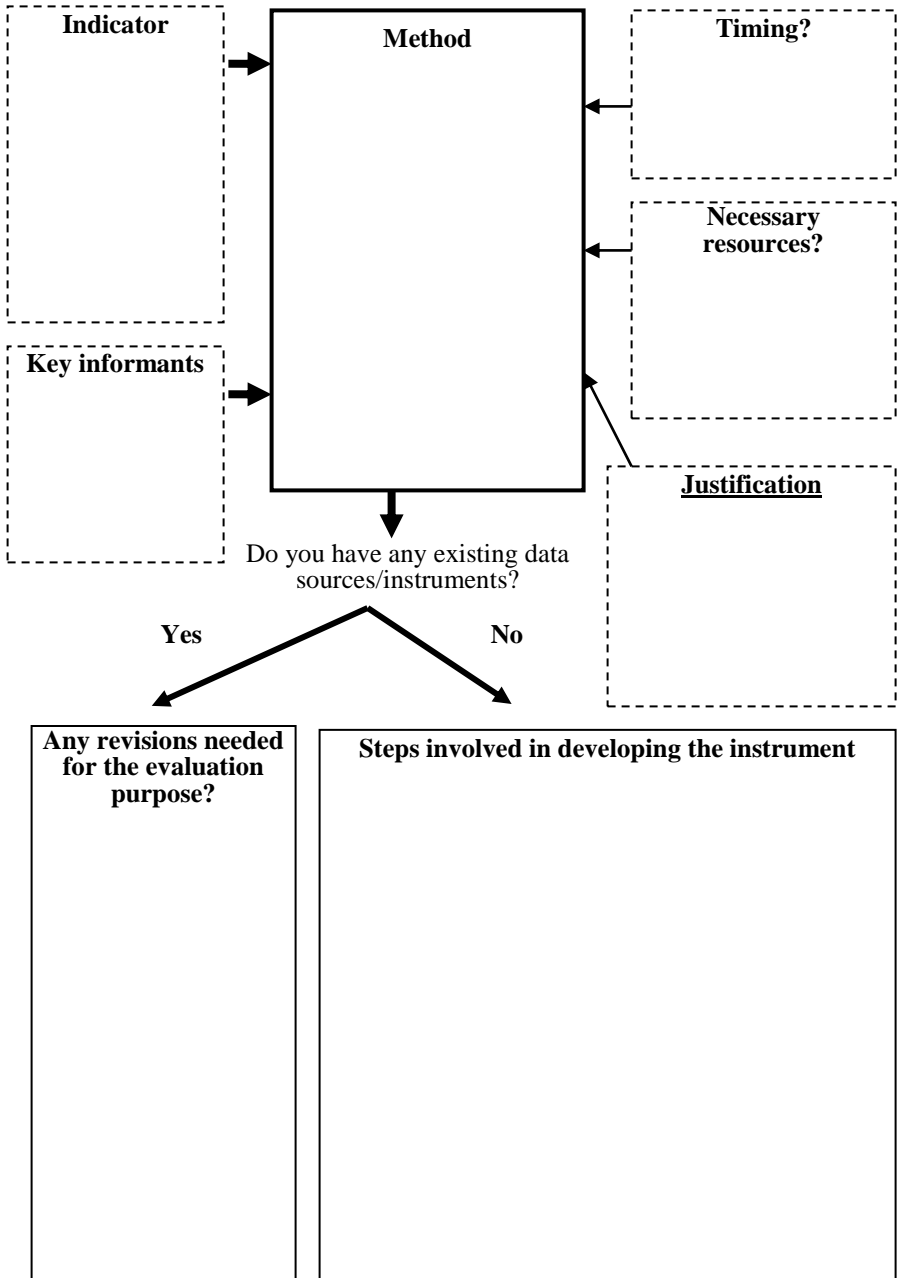
**Timing:**





**Task 3:**

- ❖ Identify the most pressing evaluation questions intended users have regarding the program, and identify what evidence will provide answers to the questions.
- ❖ For each evaluation question you identified, consider the most feasible, credible, and reliable data collection methods, in light of *who will use the findings*.



## C. Data analysis and interpretation

After collecting information, the data has to be organized in a way that is analyzable and interpretable to the primary intended users. The goal of data analysis is to provide order and structure in making sense of the information. The analysis and interpretation procedures should be informed by the evaluation questions and by trusted techniques.

### 1. Developing analysis procedures

When planning for data analysis, start by reviewing the evaluation questions. The key consideration is to *whose eyes* the analysis process and the results have to be credible and trustworthy. Take a look at the issues and considerations surrounding analysis with an example.

- ❖ **Purpose:** To inform what kind of instructional adjustments are necessary in language classes for the second half of the semester.
  - ❖ **Data collection method:** Mid-semester student focus group session (a carefully planned discussion to elicit participants' opinions and perceptions on a given topic), using outside trained facilitator to meet with the students (teacher not present).
  - ❖ **Data:** Focus group notes and recordings
- (a) Who should be involved in data analysis?
    - Program internal personnel who are not involved in teaching the students, or cooperating program external personnel.
    - Why? In order to avoid suspicion of data manipulation.
  - (b) How can data be extracted and organized?
    - Get general sense of emerging themes through repeated review of notes and audio. Extract and organize data into themes relevant to the questions.
    - Transcribing the audio-recorded data depends on how data will be used and how feasible it is to do so. Transcribe to find detailed patterns of evidence.
  - (c) What data analysis techniques will be used?
    - Identify which themes are salient; label them using informant's language.
    - Summarize general patterns and describe the range of opinions and attitudes, including disagreements and variability in opinions.
    - Extract illustrative quotes that spell out and represent themes.
  - (d) How can reliability/trustworthiness of analyses be maximized?
    - Reliability can be assured by using multiple coders (two coders for each dataset), depending on how feasible it is to have additional coders. Another strategy is to have a colleague listen to the recording and check the notes.

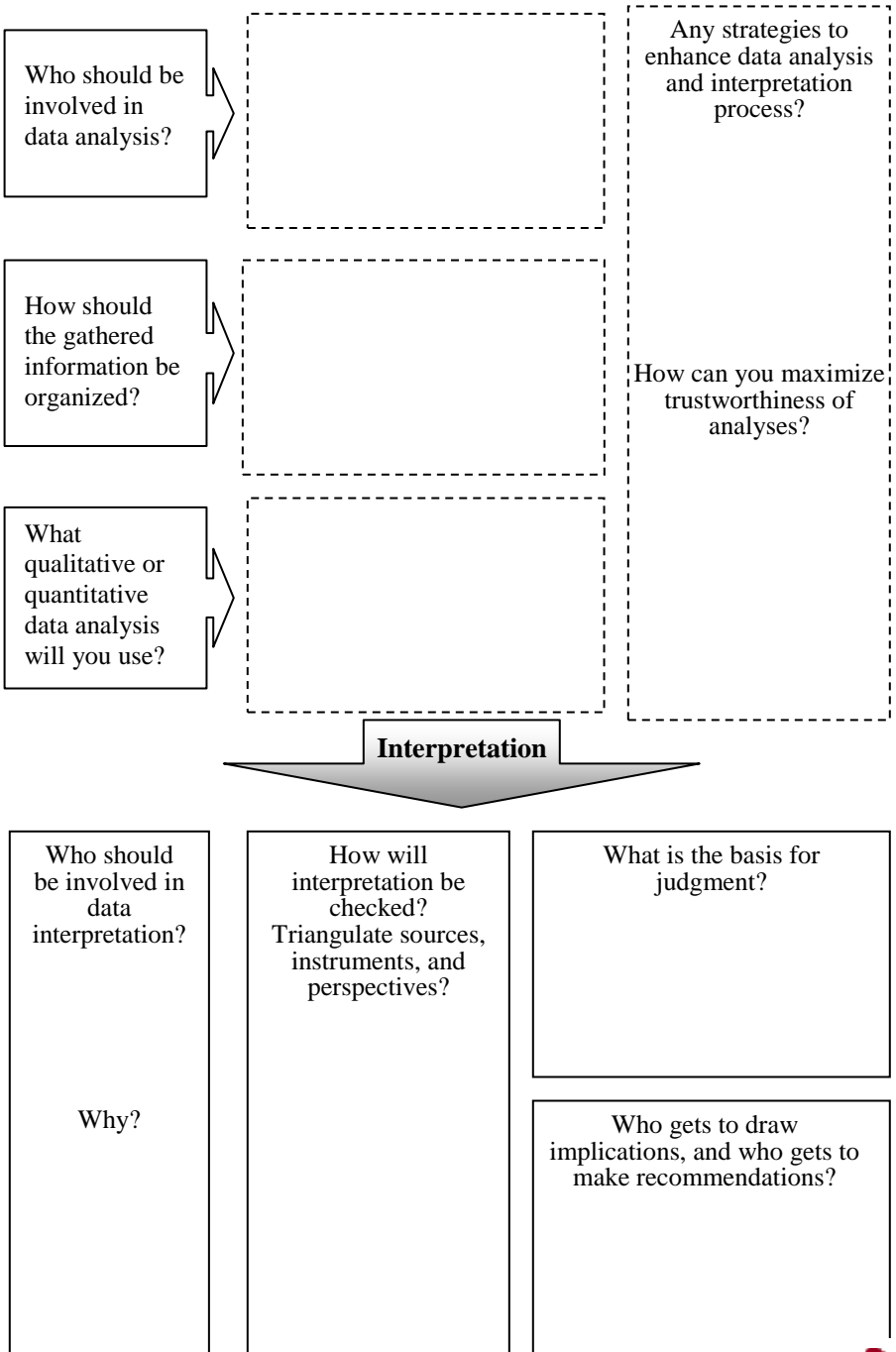
### 2. Planning for interpretation of findings

Based on the results of data analysis, interpretation and value judgments have to be made in order to respond to evaluation questions. Since interpretation is often affected by personal perspectives, careful consideration of the human factor will be necessary.

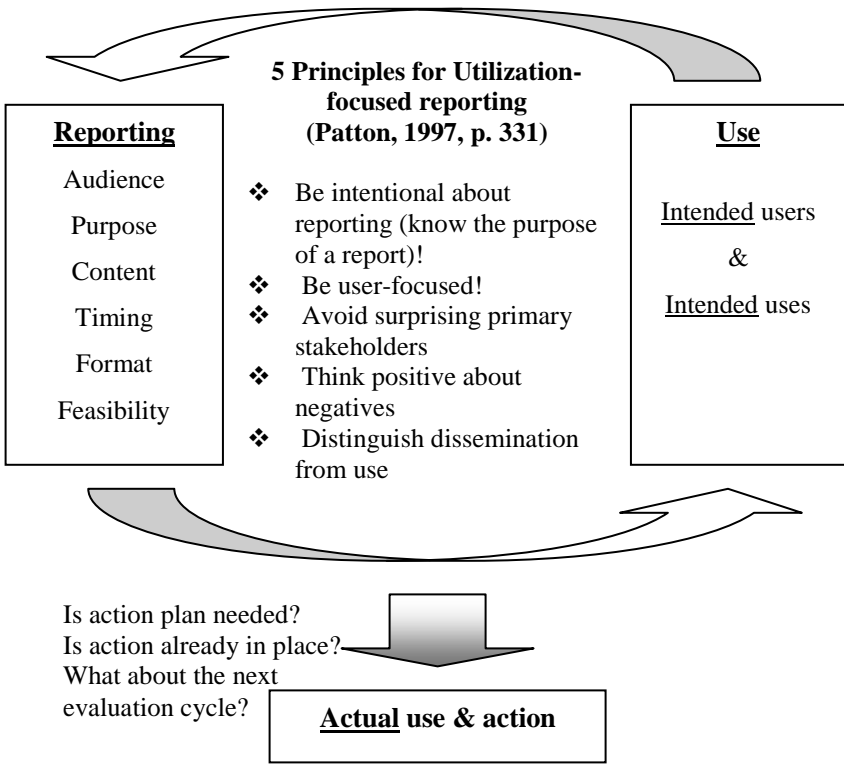
- (a) Whose perspective counts? Who should be involved in data interpretation? Why?
- (b) How will interpretation be checked? Is triangulation of sources and perspectives necessary?
- (c) What is the judgment based on? Are there any pre-set criteria for judgment?
- (d) To what extent should findings lead to implications for program change?
- (e) Who gets to draw implications, and who gets to make recommendations?
- (f) Are the interpretations and recommendations evident from the data and clearly articulated in understandable language for the intended users (and other stakeholders)?



**Task 4:** For your program evaluation, clarify data analysis and interpretation by filling in the diagram below.



## D. Reporting and using findings



### 1. Factors affecting actual use:

1. Accessibility of findings to potential users
2. Ownership of the findings
3. Increased control over evaluation
4. Development of new capacities through evaluation
5. Findings provide clear and feasible guidance for future program action.

### 2. In planning for actualization of the recommendations...

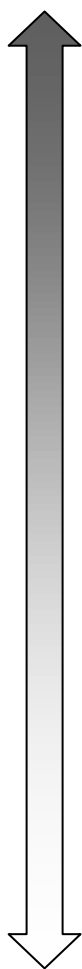
1. Review plans thus far and determine what needs to happens next.
2. Select a reasonable target date and plan backwards, considering likely time necessary for accomplishing discrete objectives.
3. Determine feasible and concrete timeline.
4. Consider who is going to be involved and at what point.

### 3. Reflective question for an evaluator:

To what extent does your plan of action incorporate other participants in the process? Can it or should it do so to a greater degree?



**Task 5:** What kind of reporting format options will work in your evaluation? Some example reporting formats are listed in the box below. Consider reporting formats from a more interaction and engagement oriented format to a more formal approach.



**More engagement and communication**

**Purposes for reporting?**

**Less engagement and more formal**

**Purposes for reporting?**

Be flexible in reporting! Consider the best format that maximizes access to and use of the findings. Here are some options for reporting formats:

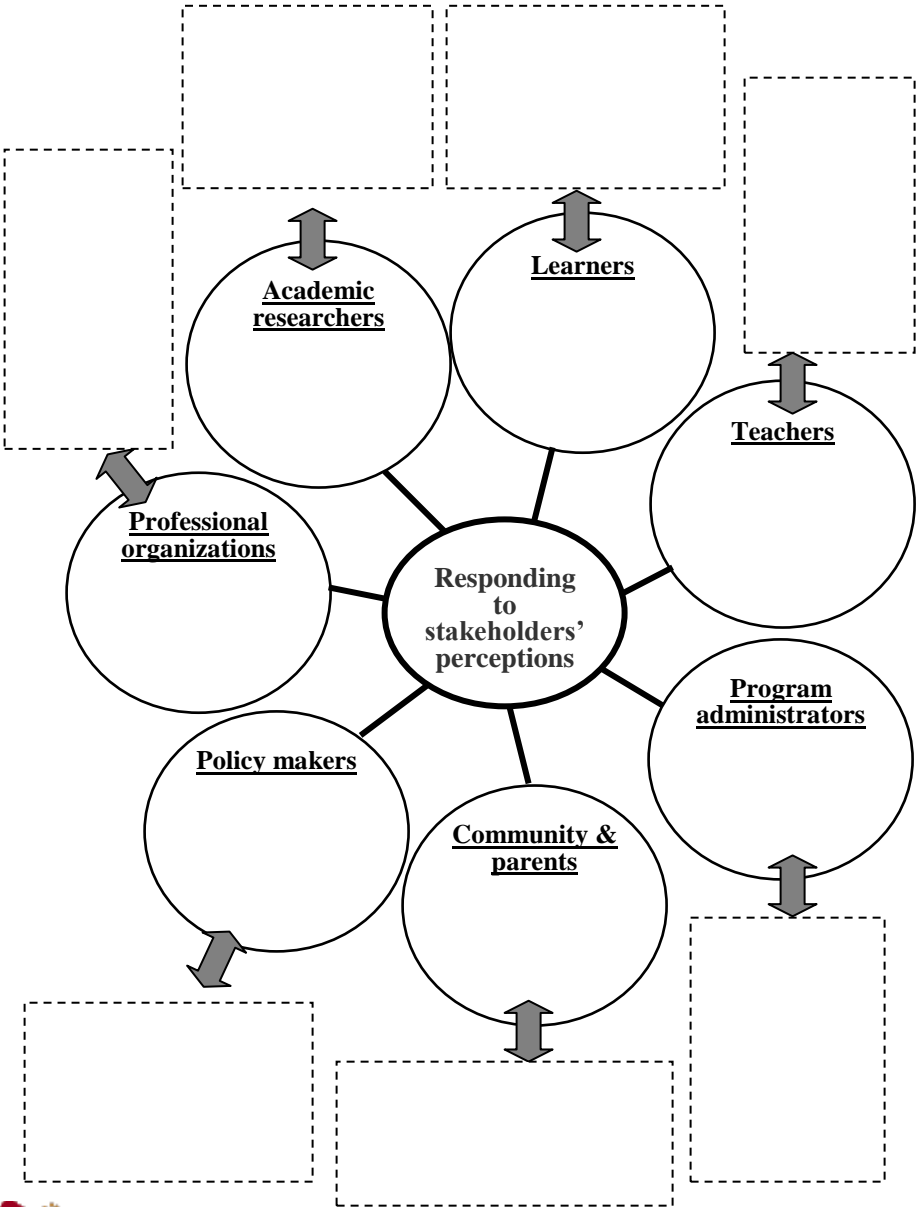
- full written report
- executive summary
- research monograph
- newsletter article
- brochure
- website, posting
- poster
- online communication
- online conference
- video-tape presentation
- online presentation
- (e.g., PPT with voice recordings)
- workshop
- oral presentation
- meeting



# E. Developing Evaluation Culture in Your Program

## Task 6:

- ❖ What perceptions (good or bad) do you think stakeholders of your program have towards program evaluation? Fill in the circles.
- ❖ In response to the perceptions, how would you explain the important roles program evaluation can play in your particular setting? How would you characterize the responsibilities of each stakeholder group? Fill in the boxes.



## References

- ACTFL (2002). Program standards for the preparation of foreign language teachers. Downloaded November 18, 2007:  
<http://www.actfl.org/files/public/ACTFLNCATEStandardsRevised713.pdf>
- Alderson, J.C., & Beretta, A. (Eds.), *Evaluating second language education*. Cambridge: Cambridge University Press.
- Faculty Working Group on Foreign Language Program Evaluation (2007). *Developing useful evaluation practices in college foreign language programs: Showcase session*. (NFLRC NetWork #48) [PDF document]. Honolulu: University of Hawai'i, National Foreign Language Resource Center.  
<http://nflrc.hawaii.edu/prodev/si07d/ADFLshowcaseBooklet.pdf>
- Fudano, H. (2005). *Nihongo kyouiku program hyoukani kannsuru kennkyuu* [An evaluation study of a Japanese education program]. Unpublished doctoral dissertation, Nagoya University, Nagoya, Japan
- Harris, J. (forthcoming). Negotiating independence and shaping debate in evaluating Irish language education. *Language Teaching Research*, 12(4).
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Newbury Park, CA: Sage.
- Llosa, L., & Slayton, J. (forthcoming). Using program evaluation to inform and improve the education of young English language learners in U.S. schools. *Language Teaching Research*, 12(4).
- Norris, J. M. (in press). *Validity evaluation in foreign language assessment*. New York: Peter Lang.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- TESOL (2002). TESOL/NCATE standards for the accreditation of initial program in P-12 ESL teacher education. Downloaded November 18, 2007:  
[http://www.tesol.org/s\\_tesol/seccss.asp?CID=219&DID=1689](http://www.tesol.org/s_tesol/seccss.asp?CID=219&DID=1689)

For extensive additional bibliographic references and related resources, check the web site of the Foreign Language Program Evaluation Project at the University of Hawaii:  
<http://www.nflrc.hawaii.edu/evaluation>



# Appendix A

## The Value of Evaluative Thinking and Action in Foreign Language Programs

A joint statement by the

*Faculty Working Group on Foreign Language Program Evaluation (2007)*

This statement is the result of discussions among the participants in the NFLRC Summer Institute 2007. The intent is to articulate the value of evaluative thinking and action to foreign language education.

Evaluative thinking and action provides a framework for discussion in programs or departments about fundamental questions of program effectiveness. These discussions can have a democratizing and unifying effect—democratizing because all voices are heard, and unifying because the process leads to communication and consensus building. Collaborative discussion and action that involves all stakeholders results in a heightened commitment of all participants to the vitality of the program, thus contributing to a sense of academic community.

The evaluation process allows faculty members to understand the program as a whole and to articulate to themselves and others what they want students to achieve in the areas of knowledge, skills, and dispositions. By identifying strengths and weaknesses, they formulate a plan (or plans) of action to increase program effectiveness and maximize student learning. The goal is to make the learning process more efficient and to create a well-articulated curriculum that is responsive to changing circumstances, all within a cyclical process of innovation and evaluation.

Evaluative thinking and action has further benefits. It enables departments to address in action-oriented ways common problems at the program level, such as low enrollments in some languages, attrition at various levels, and difficulties in the curricular transition from lower-division to upper-division courses. It offers opportunities for individual faculty members to engage in professional development activities, such as scholarship in teaching and learning and improving teaching practices through ongoing reflection. It can increase communication across departments, leading to cross-pollination between disciplines and opportunities for collaboration with colleagues on evaluation projects, as well as professional activities in other areas.

Beyond the department level, evaluative thinking and action enables faculty members to enhance the profile of their program or department within the institution by establishing themselves as leaders in evaluation initiatives and showcasing the accomplishments of their evaluation-related projects. Such leadership activities position the program or department well in requests for support (e.g., funding, faculty lines). Finally, the ability to demonstrate cycles of innovation and evaluation empowers foreign language professionals, enabling them to make a strong case for the unique contributions of language studies in a pluralist and globalized world.





# Appendix B

## Strategies for Culture Change in Program Evaluation

A joint statement by the  
*Faculty Working Group on Foreign Language Program Evaluation (2007)*

This statement is the result of discussions among participants in the NFLRC Summer Institute 2007. The intent is to encourage the foreign language field to recognize program evaluation as indispensable for enhancing student learning and program quality, and to enable the field to articulate and demonstrate—internally and externally—the unique contributions of language studies in a pluralist and globalized world.

### **Strategies for changing perceptions of evaluation and enhancing value of evaluation**

- Focus on program improvement as a goal of program evaluation.
- Emphasize the usefulness of evaluation for: (1) student learning, (2) program articulation, (3) departmental collaboration, and (4) academic community.
- Highlight the public, participatory, and inclusive nature of the evaluation process.
- Link evaluation goals to stated institutional priorities.

### **Strategies for encouraging faculty-led evaluation**

- Build on program information (curriculum, syllabi, final exams, papers, etc.) and systematize evaluation work already conducted in the department.
- Lead institutional evaluation efforts by example; forge alliances across the institution; draw on available institutional resources.
- Appropriately recognize and incentivize evaluation work within the department and the institution.
- Integrate evaluation into standard administrative, curricular, and teaching practices.
- Pursue professional development opportunities and external funding.
- Generate and showcase successful examples of evaluation.

### **Strategies for professional organizations to enhance useful evaluation**

- Recognize and disseminate successful models of program evaluation.
- Develop policy statements on useful program evaluation.
- Organize professional development events focusing on program evaluation.
- Facilitate the establishment of professional networks supporting program evaluation efforts.

N.B.: For examples of all of the above (and related resources), please monitor the Foreign Language Program Evaluation Project (FLPEP) web site:

<http://www.nflrc.hawaii.edu/evaluation>





*Foreign Language Program Evaluation Project*

University of Hawai'i at Mānoa  
Department of Second Language Studies  
1890 East-West Rd.  
Honolulu, HI 96822

Please visit our project website:  
<http://www.nflrc.hawaii.edu/evaluation>

[jnorris@hawaii.edu](mailto:jnorris@hawaii.edu)

Copyright © 2007  
John M. Norris & Yukiko Watanabe